CLAIMS

What is claimed is:

1.   A system that refines a general-purpose search engine, comprising:

a component that identifies an entry point to the general-purpose search engine; and

a tuning component that filters search query results of the general-purpose search engine based on criteria associated with the entry point.

2.   The system of claim 1, the criteria comprising one or more of a document property, a context parameter, and a configuration.

3.   The system of claim 2, the document property comprising one or more of a term that appears on a web page, a property of a Uniform Resource Locator (URL) identifying the web page, a property of a plurality of URLs that link to the web page, a property of a plurality of web pages that link to the web page, and a layout.

4.   The system of claim 2, the context parameter comprising one of a word probability and a probability distribution

5.   The system of claim 1, the tuning component provided with training data to learn what properties of a document are indicative of the document being relevant to a user executing a search query from the entry point.

6.   The system of claim 1, the tuning component configured to differentiate between a query result that is relevant to a search query context for a group of users and a query result that is non-relevant to the search query context for the group of users.

7.   The system of claim 1, the tuning component configured to employ statistical analysis in connection with filtering the search query results.

8.   The system of claim 1, the tuning component employed to generate one or more context parameters for a received query result, and then compare the generated context parameters with a relevant context parameter and a non-relevant context parameter to determine whether the query result is relevant.

9.   The system of claim 1, the tuning component further employed to rank the query results.

10.   The system of claim 9, the ranking determined by the degree of relevance of the query result to a relevant data set and a non-relevant data set, wherein the relevance is determined *via* one of a similarity measure and a confidence interval.

11.   The system of claim 9, the ranking order comprising one of ascending and descending, from the most relevant result to the least relevant result.

12.   The system of claim 1, the tuning component configured for a plurality of entry points associated with one or more groups of users.

13.   A system that tunes a general-purpose search engine, comprising:
        a filter component that parses relevant and non-relevant general-purpose search engine content results for an entry point based on training data; and
        a ranking component that sorts the filtered results in accordance with the training data for presentation to a user.

14.   The system of claim 13, the filter component parsing the results as a function of one or more of a document property, a context parameter, and a configuration associated with the entry point.

15.   The system of claim 13, the filter component trained to differentiate between a relevant and a non-relevant result *via* the training data.

16.   The method of claim 13, the training data comprising a set of relevant data associated with a search context of a user for the entry point and a et of non-relevant data comprising random data unrelated to the search context of the user for the entry point.

17.   The system of claim 13, the filter component configured to employ statistical analysis to facilitate determining whether a result is relevant or non-relevant to the entry point.

18.   The system of claim 13, the ranking component employing a technique to determine the degree of relevance of the query results with respect to a relevant data set and a non-relevant data set.

19.   The system of claim 18, the technique comprising one of a similarity measure and a confidence interval.

20.   The system of claim 13, the ranking order comprising one of ascending and descending, from the most relevant result to the least relevant result.

21.   The system of claim 18, the ranking performed on the relevant query results, wherein the non-relevant results are discarded.

22.   A method to filter and rank general-purpose search engine results associated with an entry point, comprising:

      executing a query search through the entry point;

      filtering the general-purpose search engine results; and

      ranking the general-purpose search engine results.

23.   The method of claim 22, further comprising employing a statistical hypothesis to determine whether a result is relevant or non-relevant to a search context of the entry point.

24. The method of claim 23, the statistical hypothesis employing a threshold in connection with a probability distribution for relevant data and a probability distribution for non-relevant data, wherein respective word probabilities are generated for the search query results and compared to the threshold, the probability distribution for relevant data and the probability distribution for non-relevant data to determine whether the results are relevant or non-relevant.

25. The method of claim 24, the threshold employed to bias the decision to mitigate one of a result being deemed non-relevant when the result is relevant and a result being deemed relevant when the result is non-relevant.

26. The method of claim 22, further employing a probability distribution analysis or machine learning in connection with the filtering and ranking, wherein suitable probability distributions include a Bernoulli, a binomial, a Pascal, a Poisson, an arcsine, a beta, a Cauchy, a chi-square with N degrees of freedom, an Erlang, a uniform, an exponential, a gamma, a Gaussian-univariate, a Gaussian-bivariate, a Laplace, a log-normal, a rice, a Weibull and a Rayleigh distribution, and the machine learning can classify based on one or more of a word occurrence, a distribution, a page layout, an inlink, and an outlink.

27. The method of claim 22, further comprising employing a statistical analysis to rank search query results.

28. The method of claim 27, the ranking comprising one of generating word probabilities and employing a confidence interval to determine relevance, and generating a similarity measure comprising one of a cosine distance, the Jaccard coefficient, an entropy-based measure, a divergence measure and/or a relative separation measure to determine similarity.

29.   A method to manually customize a general-purpose search engine for an entry point, comprising:

    providing a set of relevant data to train a component to discern query results relevant to a search context of a user employing the entry point; and

    providing a set of non-relevant data to train the component to discern query results unrelated to the search context, wherein the set of relevant data and the set of non-relevant data are manually provided and then employed to determine whether a query result is relevant to the search context.

30.   The method of claim 29, the set of relevant data comprising data associated with the search context of the user for the entry point.

31.   The method of claim 29, the set of non-relevant data comprising random data unrelated to the search context of the user for the entry point.

32.   The method of claim 29, further comprising providing information to associate respective query results with the entry point.

33.   The method of claim 29, the set of relevant data and the set of non-relevant data employed to train the component to learn the features that differentiate relevant data from non-relevant data.

34.   A method to automatically customize a general-purpose search engine for an entry point, comprising:

    executing a query search *via* the entry point;

    recording a query result selected by a user as relevant;

    recording a higher ranked query results, wherein a lower ranked result is selected by the user, as non-relevant; and

    providing the recorded results to automatically train the filter to discriminate between results relevant to a search context and results non-relevant to the search context.

35.   The method of claim 34, the set of relevant data comprising data associated with the search context of the user for the entry point.

36.   The method of claim 34, the set of non-relevant data comprising data unrelated to the search context of the user for the entry point.

37.   The method of claim 34, further comprising providing information to associate respective query results with the entry point.

38.   The method of claim 34, the set of relevant data and the set of non-relevant data employed to train the component to learn the features that differentiate relevant data from non-relevant data.

39.   The method of claim 34, the query results selected *via* a click thru technique, wherein a mouse is employed to select a link associated with the query result by clicking on the link.

40.   The method of claim 34, further comprising generating a word probability distribution for the relevant recorded results and a word probability distribution for the non-relevant recorded results.

41. A data packet transmitted between two or more computer components to refine a general-purpose search engine, comprising:

a component that accept search query results for a group of users, a component that identifies one or more entry points associated with the search, a component that employs a relevant data set and a non-relevant data set to determine whether a search result is relevant, and a component that ranks the search results based on the degree of relevance to the group of users and the entry point.

42.   A computer readable medium storing computer executable components that tunes a general-purpose search engine to improve context search query results, comprising:

a component that filters the general-purpose search engine results based on training data sets; and

a component that ranks the general-purpose search engine results according to the similarity of the search engine results to the training data sets.

43.   A system that filters and ranks general-purpose search engine results, comprising:

means for filtering general-purpose search engine results to determine whether a query result is relevant to a search context of a group of users and an entry point, and

means for ranking the general-purpose search engine results based on a relevance of the general-purpose search engine results to the search context of a group of users and an entry point.